Science & Technology Facilities Council

SEVENTH FRAMEWORK PROGRAMME

metafor

"finding and understanding simulations of past, present and future climate"

Model Metadata for CMIP5

Presented by

B. Lawrence (BADC)

On behalf of the Metafor and Earth System Curator teams

Earth System Curator

Spanning the Gap Between Models and Datasets

Science & Technology Facilities Council

Outline

- Context
  - The METAFOR Project
  - ESG Curator
- Documenting Models and Simulations
  - The METAFOR Common Information Model (CIM)
  - Controlled Vocabulary
- The “Questionnaire” for CMIP5
  - Model Descriptions
    - Component Structure
    - Component Attributes
  - Experiments Simulations
- Using the model metadata
- Summary

Science & Technology Facilities Council

Discrimination and Documentation "A problem for CMIP5"

CMIP5 will involve more than twenty different models running at least twenty institutions running dozens of simulations corresponding to dozens of numerical experiments.

The volume of data will be petabytes. Even the core data will be of order ¾ of a petabyte.

What should one download?

What are the chacteristics of the model that produced it?

How was a particular simulation constructed?

Rowan Sutton: "There are no end-users of climate predictions"

Gavin Schmidt: "the public … will get used to dealing with climate model outputs … however, an increased amount of hand-holding will be necessary"

But engagement requires effort and time!

... and it's not always the public who require "engagement" above and beyond "normal academic behaviour", misunderstanding abound and even appear in the peer reviewed literature!

Science & Technology Facilities Council

Metafor: Facts and Figures

Common Metadata for Climate Modelling Digital Repositories

http://metaforclimate.eu

SEVENTH FRAMEWORK PROGRAMME

CAPACITIES

INFRA-2007-1.2.1 - Scientific Digital Repositories

Heritage: ENES:

- PRISM project (FP5 2001-2004)
- Followed by the PRISM Sustained Initiative (PSI)

11 partners

Total budget of 2.2M€

Started March 2008, duration 3 years

- NCAS Climate, NCAS CMS. Reading Uni, UK (Coordinator)
- NCAS BADC (STFC CEDA), UK
- CERFACS, France
- Models and Data, Max Planck Institute for Meteorology, Germany
- Institute Pierre-Simon Laplace, CNRS, France
- University of Manchester, UK
- Met Office, UK
- Administratia Nationala de Meterologie, Romania
- Météo France, CNRM, France
- CLIMPACT, France
- CICS, Princeton University, USA

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

CERFACS

M&D Modèle & Données

Sciences de l'environnement

MANCHESTER 1824

Met Office

METEO FRANCE

climpact
CLIMATE INFORMATION TECHNOLOGY

CICS
COOPERATIVE INSTITUTE FOR CLIMATE SCIENCE

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Context: Earth System Curator

GOALS The Curator project develops tools and portals that link simulations, models, components, datasets, and services in support of climate research and assessments. Curator workflows are driven by standardized metadata developed in collaboration with the Earth System Grid (ESG) and EU metafor projects.

PARTNERS NOAA GFDL, NCAR, Georgia Institute of Technology, Massachusetts Institute of Technology, plus external partners

External Partners

Earth System Grid

ESMF

PROJECTS

- **Metadata-enhanced science gateway** (lead NCAR)
  Enhance the ESG portal with model metadata for improved understanding of datasets and their source models
- **Automatic code generation** (lead Georgia Tech)
  Generate component couplers automatically using ESMF and the Eclipse Modeling Framework
- **Climate modeling workflow** (lead GFDL)
  Create a Flexible Run-Time Environment (FRE) that enables modelers to configure, execute, archive, reproduce, and modify climate model runs

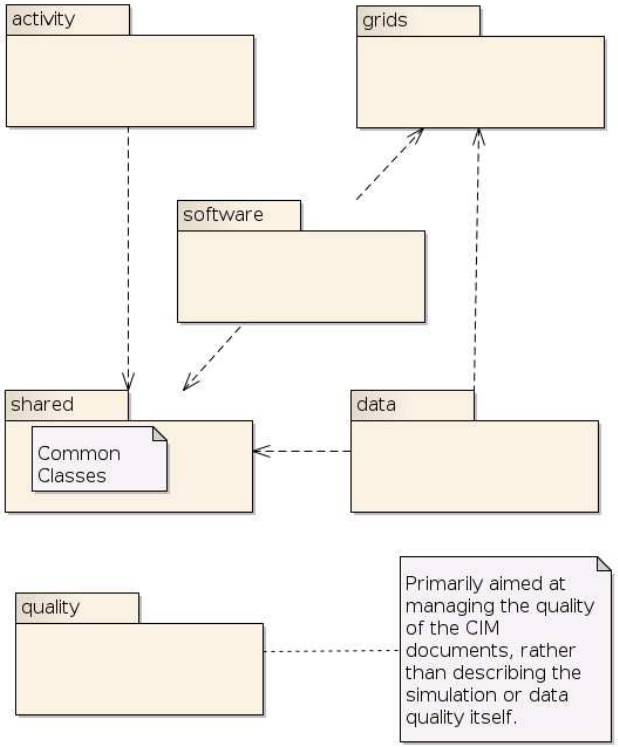
Earth System Curator

Spanning the Gap Between Models and Datasets

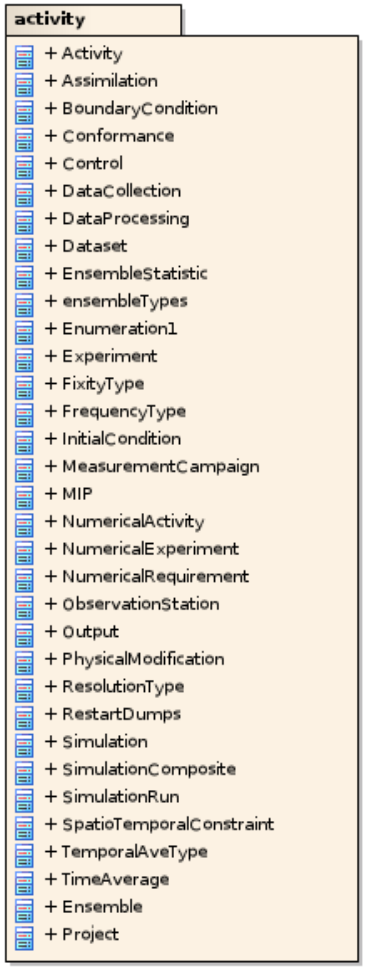
Slide Courtesy of Cecelia DeLuca , NCAR

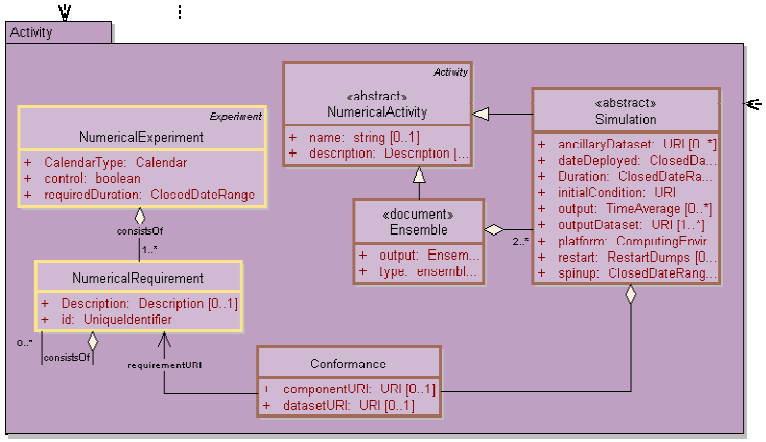
Curator is funded by the National Science Foundation and NASA

The Metafor Common Information Model



An activity uses software to produce data to be archived in a repository.



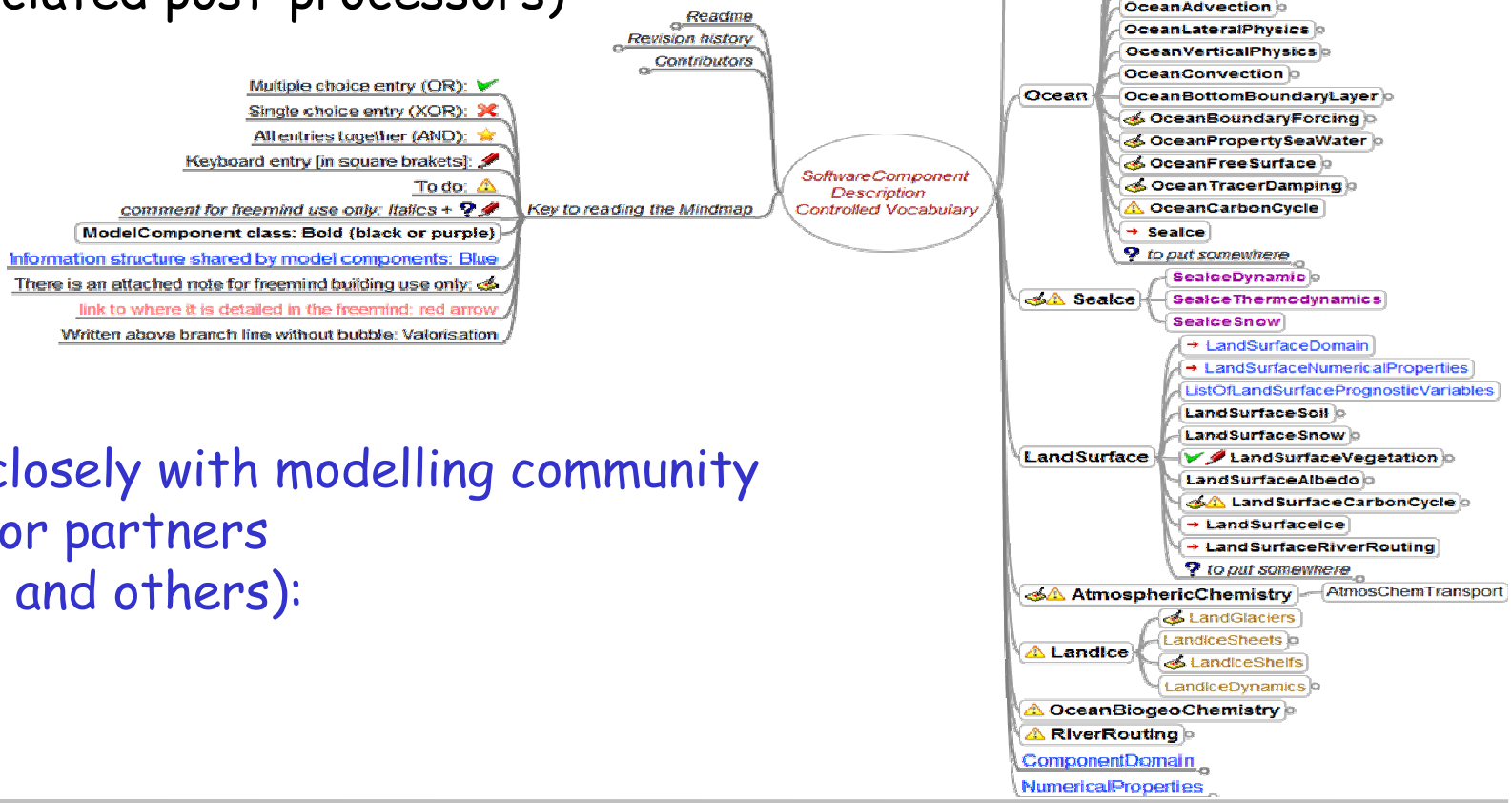


An simulation is a numerical activity which conforms to the numerical requirements of a numerical experiment.

Six packages with complex structure and various levels of interdependency.

Science & Technology Facilities Council

Metafor developing controlled vocabularies to describe models (and associated post-processors)



Working closely with modelling community and Curator partners (via Balaji and others):

Science & Technology Facilities Council

CMIP5 and Model Metadata

CMIP5:

> 20 modelling centres

> 50 numerical experiments

> 86 simulations (total ensemble members) within experiments

> 6500 years of simulation

> Data to be available from “core-nodes” and “modelling-nodes” in a global federation.

> Users need to find datasets, and discriminate between models, and between simulation characteristics.

CMIP5 Metadata Requirements:

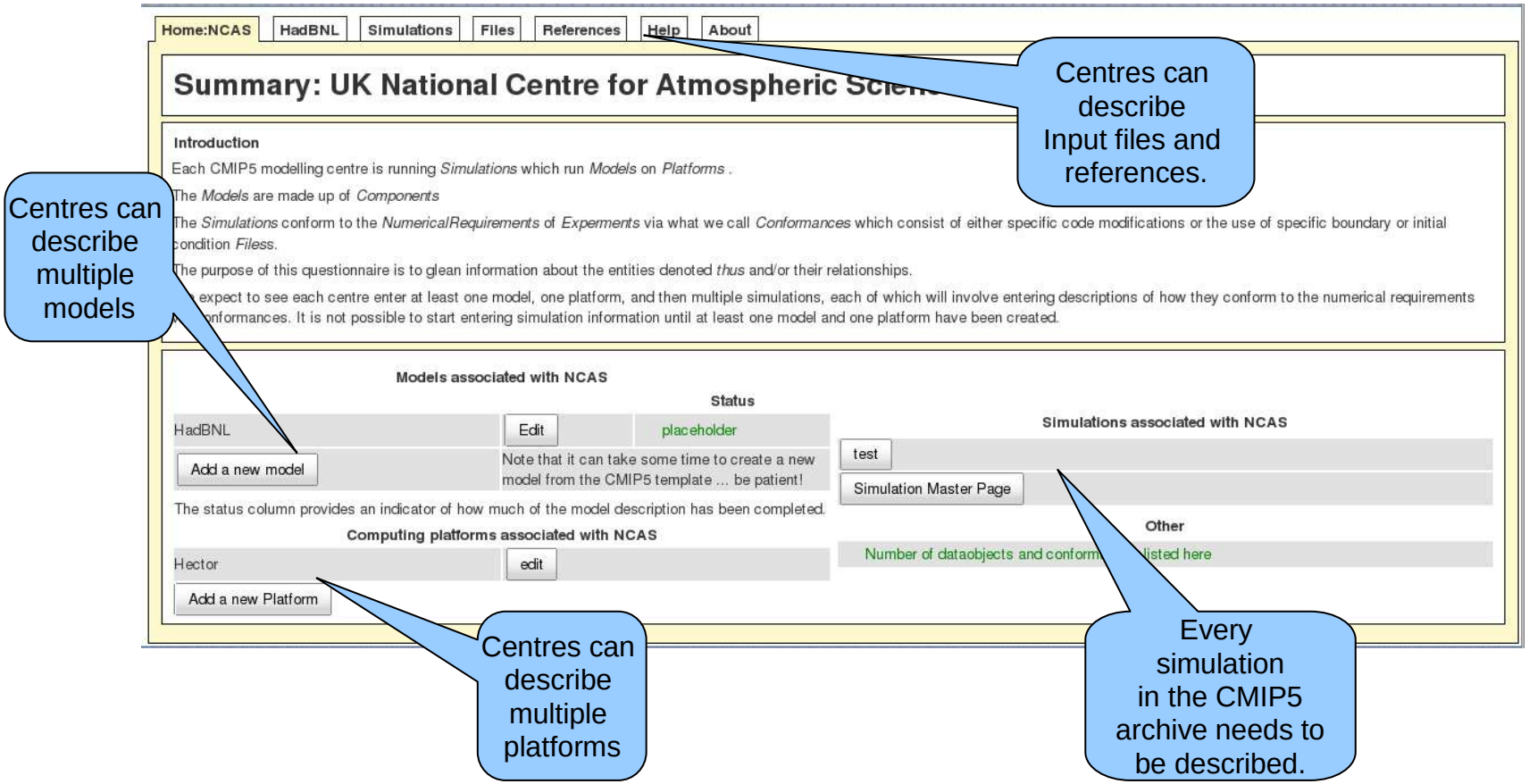
> Internal file metadata as required by CMOR.

> Grid definitions using gridspec files.

> Model Metadata as required by the “official” model metadata questionnaire ...

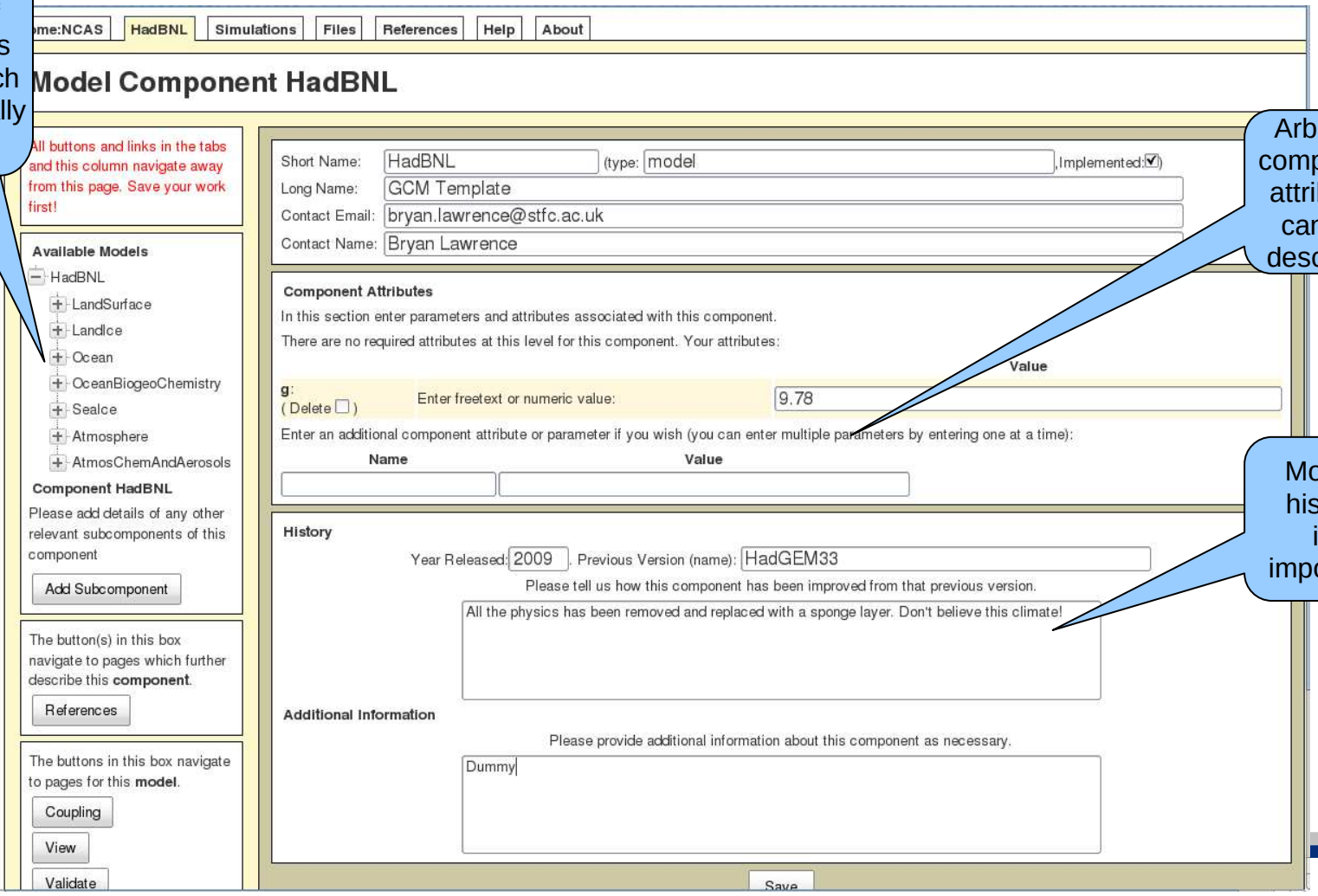
> …which is under active development. The current DRAFT version is available at http://cmip5.metafor.ceda.ac.uk (don't enter any real information in this version, it is being updated weekly and all data is deleted).

Questionnaire Screenshots, 1: Summary Page



Questionnaire Screenshots 2: Model Description

Every model consists of components each of which are individually described

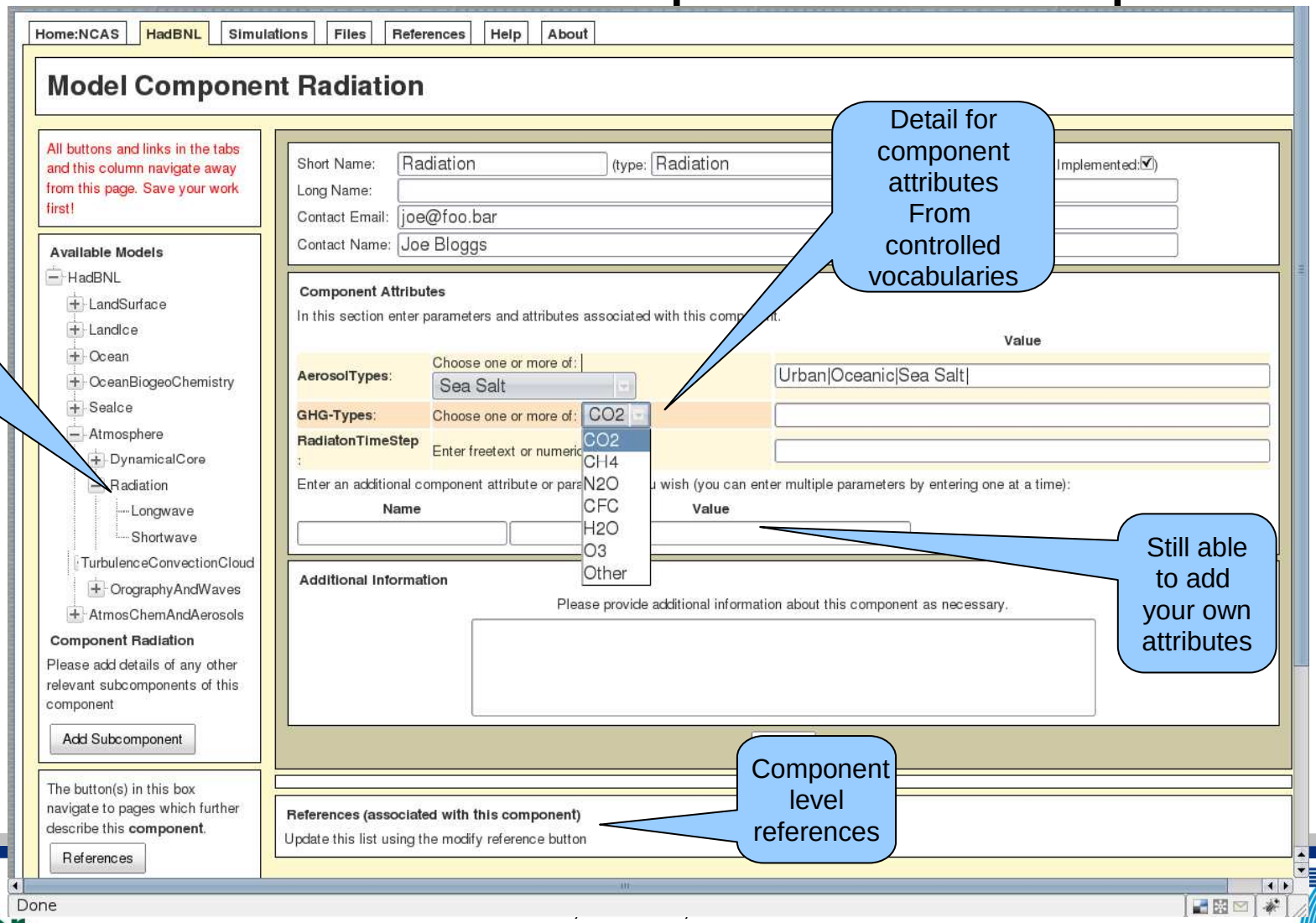


Arbitrary component attributes can be described

Model history is important

Science & Technology Facilities Council

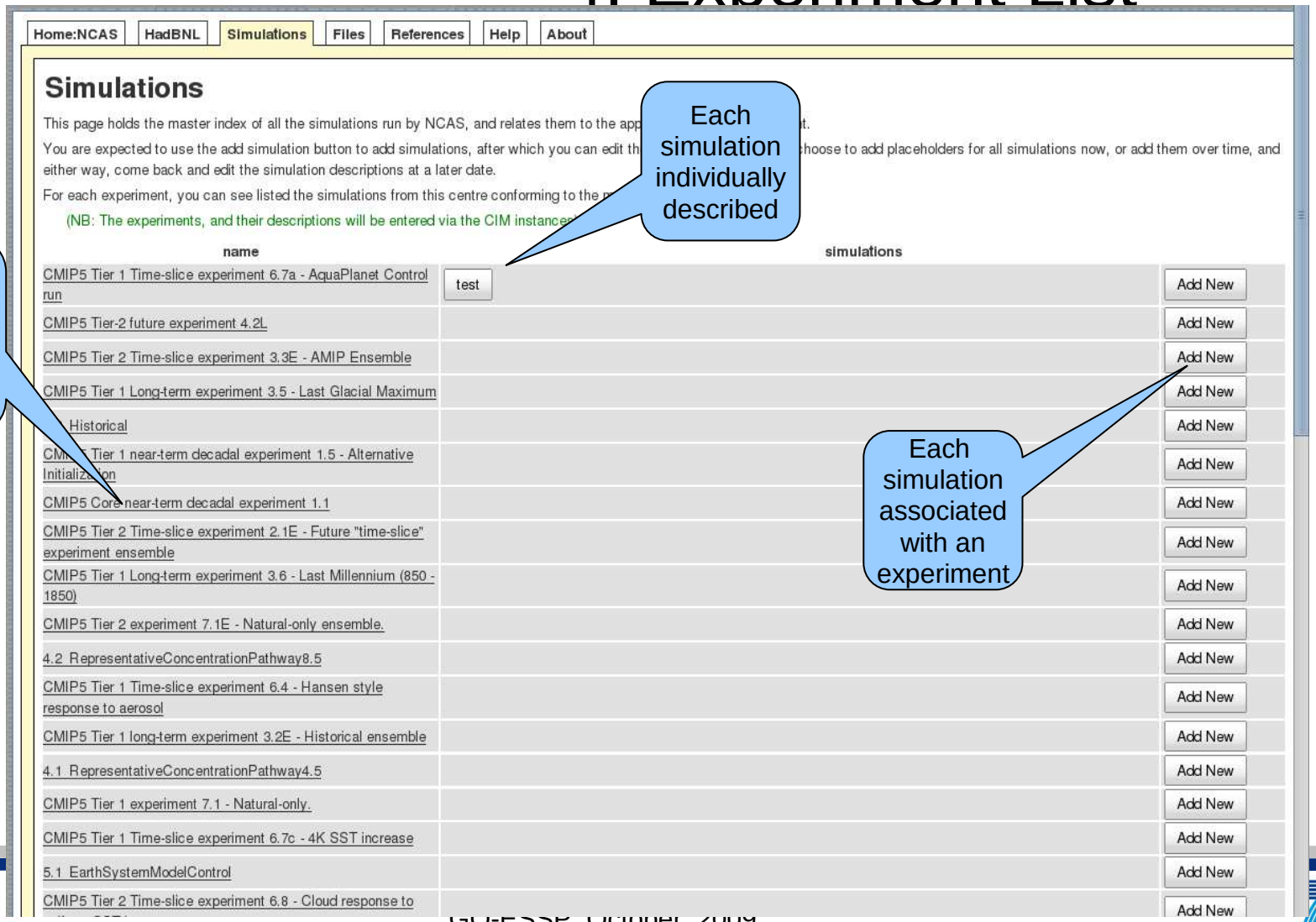
Questionnaire Screenshots 3: Component Description



metafor

SEVENTH FRAMEWORK PROGRAMME

Questionnaire Screenshots 4: Experiment List

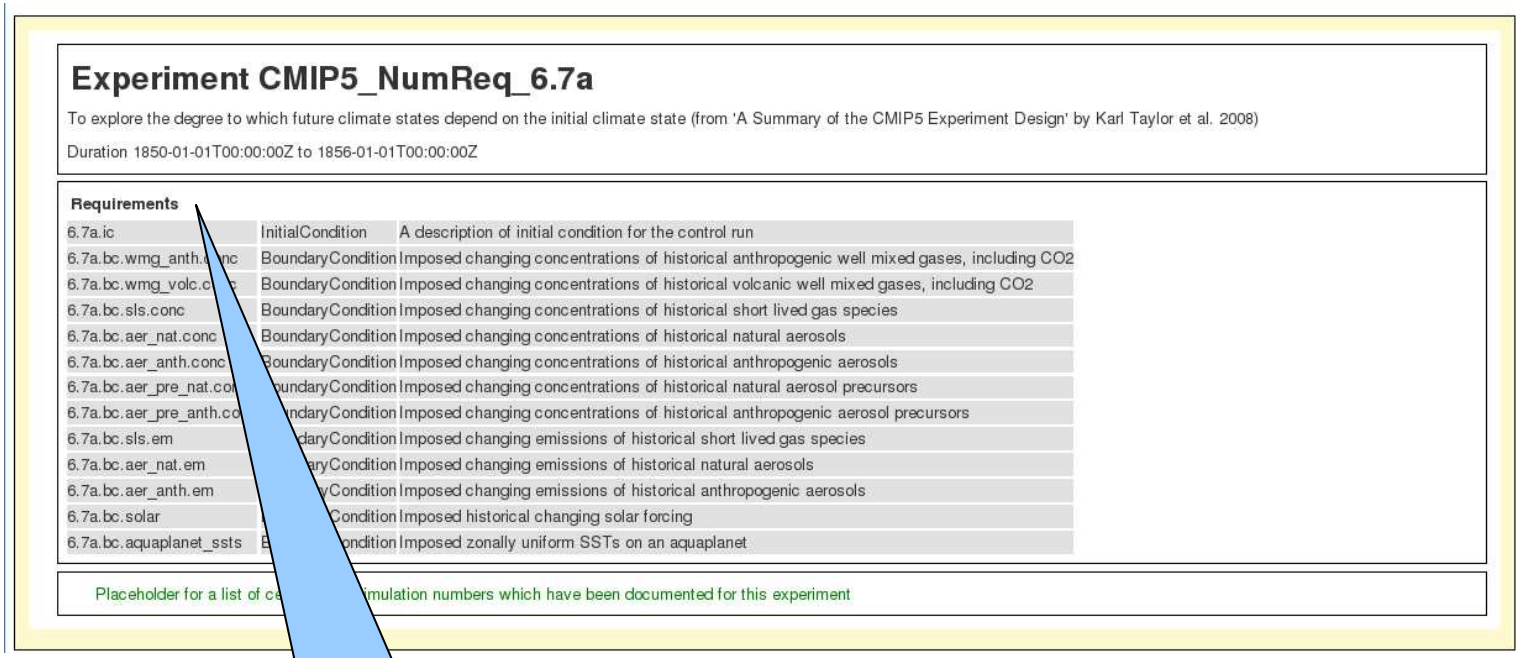


Each simulation individually described

All CMIP5 experiments fully described (see next slide)

Each simulation associated with an experiment

Questionnaire Screenshots 5: Experiment Description



Experiment CMIP5_NumReq_6.7a

To explore the degree to which future climate states depend on the initial climate state (from 'A Summary of the CMIP5 Experiment Design' by Karl Taylor et al. 2008)

Duration 1850-01-01T00:00:00Z to 1856-01-01T00:00:00Z

Requirements

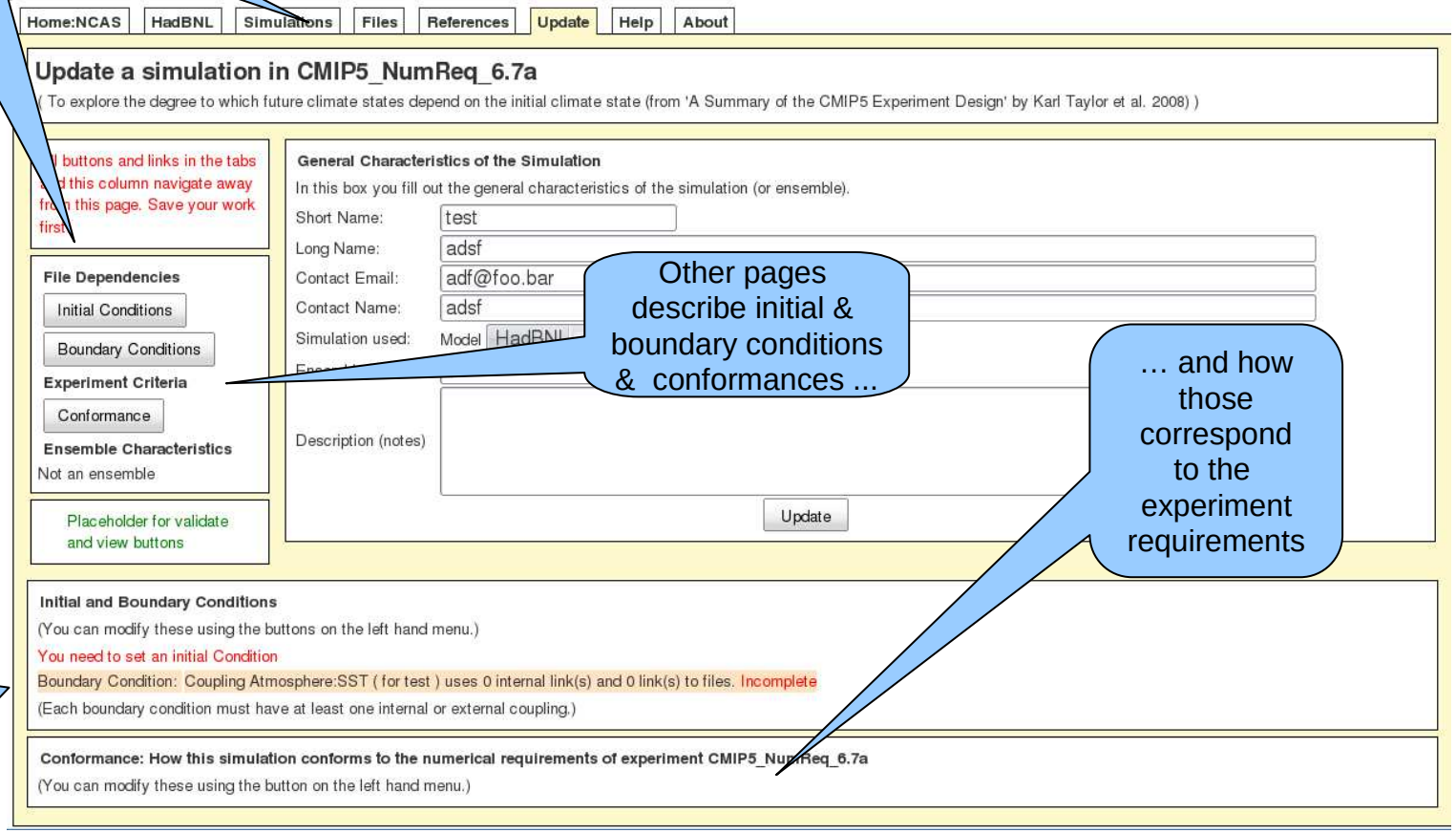
6.7a.ic	InitialCondition	A description of initial condition for the control run
6.7a.bc.wmg_anth.conc	BoundaryCondition	Imposed changing concentrations of historical anthropogenic well mixed gases, including CO2
6.7a.bc.wmg_volc.conc	BoundaryCondition	Imposed changing concentrations of historical volcanic well mixed gases, including CO2
6.7a.bc.sls.conc	BoundaryCondition	Imposed changing concentrations of historical short lived gas species
6.7a.bc.aer_nat.conc	BoundaryCondition	Imposed changing concentrations of historical natural aerosols
6.7a.bc.aer_anth.conc	BoundaryCondition	Imposed changing concentrations of historical anthropogenic aerosols
6.7a.bc.aer_pre_nat.conc	BoundaryCondition	Imposed changing concentrations of historical natural aerosol precursors
6.7a.bc.aer_pre_anth.conc	BoundaryCondition	Imposed changing concentrations of historical anthropogenic aerosol precursors
6.7a.bc.sls.em	BoundaryCondition	Imposed changing emissions of historical short lived gas species
6.7a.bc.aer_nat.em	BoundaryCondition	Imposed changing emissions of historical natural aerosols
6.7a.bc.aer_anth.em	BoundaryCondition	Imposed changing emissions of historical anthropogenic aerosols
6.7a.bc.solar	BoundaryCondition	Imposed historical changing solar forcing
6.7a.bc.aquaplanet_ssts	BoundaryCondition	Imposed zonally uniform SSTs on an aquaplanet

Placeholder for a list of ca… …mulation numbers which have been documented for this experiment

Requirements for initial and boundary conditions preloaded in questionnaire setup.

Questionnaire Screenshots 6: Simulation Description

Page layout (tabs & buttons) will change before release



Other pages describe initial & boundary conditions & conformances ...

… and how those correspond to the experiment requirements

Need to *complete* Initial and Boundary conditions

Science & Technology Facilities Council

Summary

- CMIP5 Model metadata requirement in response to to large international projects capturing details of what is needed.
- Considerable flexibilty in what can be captured, but much is required too.
- The “questionnaire tool” to collect the metadata will change considerably in the near future, but will be available by late (NH) autumn.
- The archive centres intend to mandate the requirement for this metadata before making data available to community.
- ESG and Metafor/IS-ENES will deploy catalogues to exploit this metadata, linking data to information about the data allowing a wider community to accurately exploit the considerable investment in CMIP5 design and execution.